Homework 8

Charles Swarts
Swarts2

March 2016

1 8.14

1.1 question

A Queue

A bus is supposed to arrive at a bus stop every hour for 10 hours each day. The number of people who arrive to queue at the bus stop each hour has a Poisson distribution, with intensity 4. If the bus stops, everyone gets on the bus and the number of people in the queue becomes zero. However, with probability 0.1 the bus driver decides not to stop, in which case people decide to wait. If the queue is ever longer than 15, the waiting passengers will riot (and then immediately get dragged off by the police, so the queue length goes down to zero). What is the expected time between riots?

Solution: I'm not sure whether one could come up with a closed form solution to this problem. A simulation is completely straightforward to write. I get a mean time of 441 hours between riots, with a standard deviation of 391. It's interesting to play around with the parameters of this problem; a less conscientious bus driver, or a higher intensity arrival distribution, lead to much more regular riots.

1.2 raw simulation code

```
    days_Simulated <- 1000000

intensity <- 4
max_Line_Length <- 15
prob_busDriver_flakes <- .1
timelength <- 1

#The strategy is to make a sample space for
#arrival rate of the people and the arrival
#of the bus driver. In simulating the line,
#I will make a sample space that includes
#probabilities for 0:max_Line_Length,
#(>max_line_length), because if that many
#people show up, the line automatically riots.
```

```
sampleSpace <- 0:(max_Line_Length+1)
sampleSpaceProbs <- c()
for(i in 0:max_Line_Length){
  sampleSpaceProbs <- c(sampleSpaceProbs,((intensity*timelength)^i)/(factorial(i))*exp((-1)*intensity*
}
sampleSpaceProbs <- c(sampleSpaceProbs, 1-sum(sampleSpaceProbs))


arrivals <- sample(sampleSpace,10*days_Simulated,replace=TRUE,prob=sampleSpaceProbs)


busSpace <- c(0,1)
busSpaceProbability <- c(.1,.9)


busComes <- sample(busSpace,10*days_Simulated,replace=TRUE,prob=busSpaceProbability)

peopleInLine <- 0
#riots is a list of time-lengths between riots.
riots <- c()
counter <- 1
#The riot Counter is the number of hours between riots.
riotCounter <- 0
for(i in 1:(days_Simulated*10)){
  riotCounter <- riotCounter+1
  peopleInLine <- peopleInLine+arrivals[i]

  if(peopleInLine>max_Line_Length){
    riots <- c(riots,riotCounter)
    riotCounter <- 0
    peopleInLine <- 0
    }
  if(busComes[i]==1){peopleInLine <- 0}

  #This code follows from the fact that the bus
  #runs 10 times/day and at night, people who
  #are waiting go home.
  if((counter%%10)==0) {peopleInLine <- 0}
  counter <- counter+1;
}

mean(riots)
sd(riots)
```

1.3 Answers

Simulating 1000000 days,

I got the mean time between riots was 491.7139 hours. And the Standard Deviation between time between riots was 492.683.

2 8.15

2.1 question

Inventory

A store needs to control its stock of an item. It can order stocks on Fri- day evenings, which will be delivered on Monday mornings. The store is old- fashioned, and open only on weekdays. On each weekday, a random number of customers comes in to buy the item. This number has a Poisson distribution, with intensity 4. If the item is present, the customer buys it, and the store makes 100; $otherwise, thecustomerleaves.Eacheveningatclosing, thestoreloses$10 for each unsold item on its shelves. The store's supplier insists that it order a fixed number k of items (i.e. the store must order k items each week). The store opens on a Monday with 20 items on the shelf. What k should the store use to maximise profits?

Solution: I'm not sure whether one could come up with a closed form solution to this problem, either. A simulation is completely straightforward to write. To choose k, you run the simulation with different k values to see what hap- pens. I computed accumulated profits over 100 weeks for different k values, then ran the simulation five times to see which k was predicted. Results were 21, 19, 23, 20, 21. I'd choose 21 based on this information.

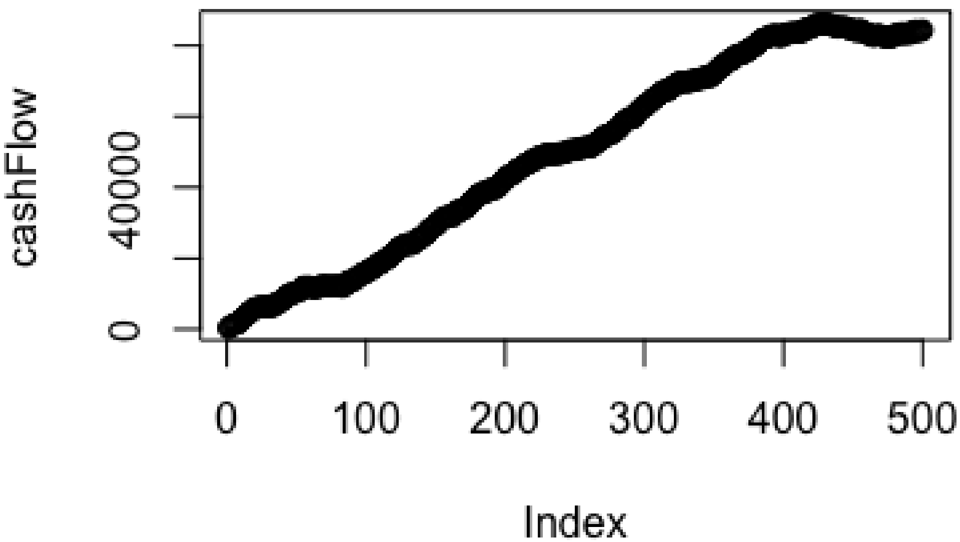
2.2 raw simulation code

```
    weeks_Simulated <- 100

original_stock <- 20
intensity <- 4
k <- 19

#My strategy for calculating the probabilities
#of the sample space for the number of
#customers who walk through the is to caluculate
#up to 2 * Expected[people who come in during the week]
#because I think it would be unlikely that that
#number of people walks into the store.
#The expected # of people for a poisson dist
#is intensity * time or 5*4 which is 20
#Therefore I will be calculating up to 40.

#numPeople()

prob_Num_People <- c()

for(i in 0:(intensity*5*2)){
prob_Num_People <- c(prob_Num_People,((intensity)^i)/(factorial(i))*exp((-1)*intensity))
}
prob_Num_People <- c(prob_Num_People,1-sum(prob_Num_People))

Num_People <- 0:(intensity*5*2+1)

people_by_Day <- sample(Num_People,weeks_Simulated*5,replace=TRUE,prob=prob_Num_People)
```

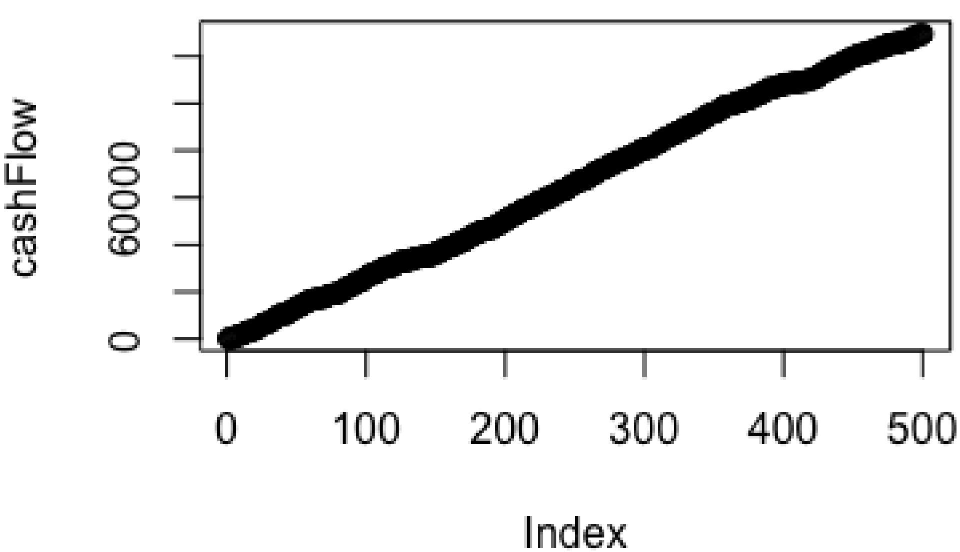
```
#people_by_Day

cashFlow <- c(0)

for(i in 1:(weeks_Simulated*5)){
  if(i!=1){
  if(people_by_Day[i]>original_stock)
  {
    cashFlow[i] <- (cashFlow[i-1]+100*original_stock)
    original_stock <- 0
  }
  else{
    cashFlow[i] <- (cashFlow[i-1]+100*people_by_Day[i])
    original_stock <- original_stock-people_by_Day[i]
  }
  }
  else{
    if(people_by_Day[i]>original_stock)
    {
      cashFlow[i] <- (100*original_stock)
      original_stock <- 0
    }
    else{
      cashFlow[i] <- (100*people_by_Day[i])
      original_stock <- original_stock-people_by_Day[i]
    }
  }
  cashFlow[i] <- (cashFlow[i]-10*original_stock)
  if(i%%5==0){original_stock <- original_stock+k}
}
plot(cashFlow)
people_by_Day
cashFlow[length(cashFlow)]
```

2.3 Answers

I did not throw product out at the end of the week, so when I plotted what it looks like when you order 20, I got plots like this:



where I only made $84400 after 100 weeks, and plots like this when I ordered 19 each week



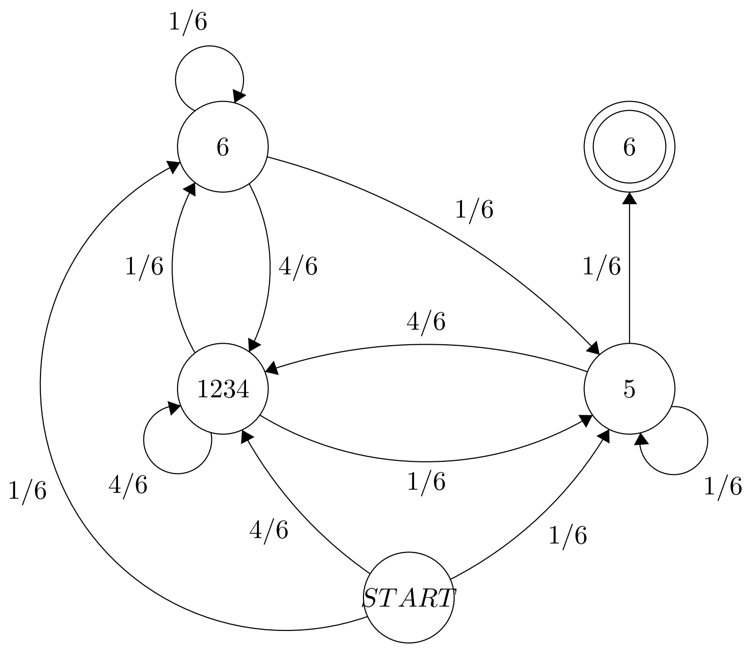
and made $129550
therefore I believe 19 to be the best amount to order each week.

3 8.2

3.1 question

8.2. Multiple die rolls: You roll a fair die until you see a 5, then a 6; after that, you stop. Write P(N) for the probability that you roll the die N times.
(a) What is $P(1)$?
(b) Show that $P(2) = (1/36)$.
(c) Draw a finite state machine encoding all the sequences of die rolls that you could encounter. Don't write the events on the edges; instead, write their probabilities. There are 5 ways not to get a 5, but only one probability, so this simplifies the drawing.
(d) Show that $P(3) = (1/36)$.
(e) Now use your finite state machine to argue that $P(N) = (5/6)P(N1) + (25/36)P(N2)$.

3.2 answer



(a). $P(1)$ according to the diagram is 0 because there is no walk of length one from START to 6. Alternative reasoning is that you can't get a five then a 6 in one role.

(b). $P(2)$ according to the diagram is 1/36 because there is only one walk of length 2 from START to 6 and one of the edges has a $P(E) = 1/6$ and the other has $P(E) = 1/6$ so to traverse both is $P(E) \cap P(E)$ which is P(E)*P(E) since they are independent, and $1/6 * 1/6 = 1/36$ Alternative reasoning is that you would have to get a 5 and then a 6 to end the game and the probability of getting a 5 then a 6 is $1/6 * 1/6 = 1/36$
(c) See diagram above
(d) $P(3)$ according to the diagram is 1/36 because the three walks that go from start to three are

$$START \to 5 \to 5 \to 6 = 1/6 * 1/6 * 1/6 = 1/216$$

$$START \to 6 \to 5 \to 6 = 1/6 * 1/6 * 1/6 = 1/216$$

$$START \to 1234 \to 5 \to 6 = 1/6 * 1/6 * 1/6 = 4/216$$

adding all the possibilities together makes $6/216 = 1/36$ This follows the general formula

$$P(n) = \sum_j P((n-1)|j)$$

Alternative reasoning is that in order to roll three before the game is over, you would have to get a 5 and a 6 on the second and third rolls respectively. It doesn't really matter the outcome of the first state, so you can say it could be "anything," and "anything" has a probability of 1. $P(2)$ has a probability of $1/36$ so $1 * 1/36 = 1/36$ which also begets the answer.

4 8.3

4.1 question

8.3. More complicated multiple coin flips: You flip a fair coin until you see either HTH orTHT, and then you stop. We will compute a recurrence relation for $P(N)$.

(b) Write $\sum_N$ for some string of length N accepted by this finite state machine. Use this finite state machine to argue that $\sum_N$ has one of four forms:

$$\mathbf{1.}TT \sum_{N-2}$$

$$\mathbf{2.}HH \sum_{N-2}$$

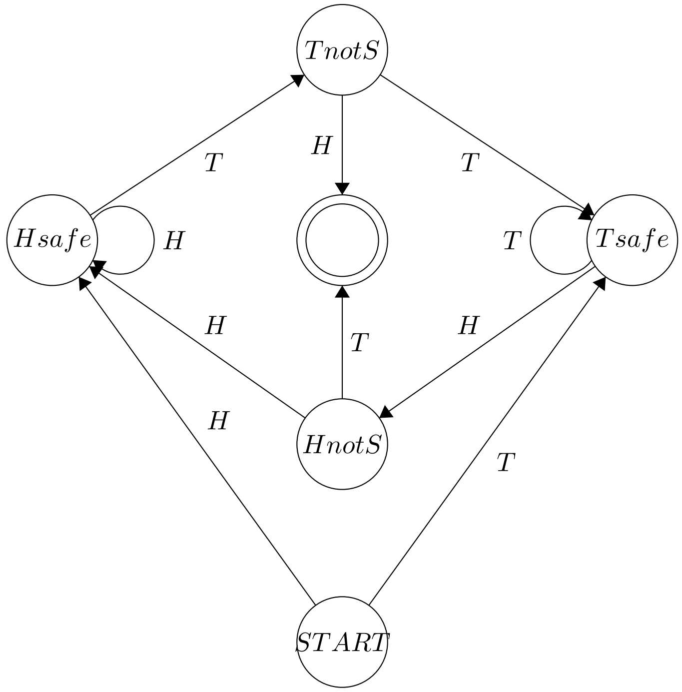
$$\mathbf{3.}THH \sum_{N-3}$$

$$\mathbf{4.}HTT \sum_{N-3}$$

(c) Now use this argument to show that

$$P(N) = (1/2)P(N-2) + (1/4)P(N-3).$$

4.2 answer



(b) So the triggers for the end state of this finite state machine are THT and HTH. And also I define the word "safe" to mean: can't cause the end state in one move. At the beginning, after the first coin flip you are safe, because if you draw a heads or a tail on the second flip, you still haven't triggered the end state.

Let's start with base cases, either the first is H or T, if it's H then, according to the state machine, the next sequence has to be either H, or TT for the sequence to be safe again. If you start out with T, then the next sequence has to be either T or HH before you know you are safe again.

Being safe means that you have recreated the conditions at the start for the last flip in the running sequence. IE you get the HTT scenario, it has now become like if Tails was the first outcome you got so that the two sequences availible to you are T and HH

It must be noted here that the $\sum_N$ implies that it is based off of what came before it. This is so you can never get the case $TTHTTH \sum N - 6$

Also, for each flip, the rest of the sequence decreases by one.

So to conclude all of this, there are four starting sequences that result,

$$1 : HH \sum_{N-2}$$

$$2 : TT \sum_{N-2}$$

$$3 : HTT \sum_{N-3}$$

$$4 : THH \sum_{n-3}$$

(c) So if we have the previous as the possible outcomes, we can break them down into their component parts which are:

1: The initial outcome. The initial outcome is boring because it is always safe, so the probability of it being correct is 1.

2:The second component which gives a 1/2 probability that the sequence is determined by the $P(N-2)$ or that it will go to the alternative.

3: In the alternative, we once again have a 1/2 chance that the variable lines up and the Probability is based on $P(N-3)$

That gives an overall probability for $P(N)$

$$P(N) = 1 * \frac{1}{2} * P(N-2) + 1 * \frac{1}{2} * \frac{1}{2} * P(N-3) = \frac{1}{2}P(N-2) + \frac{1}{4}P(N-3)$$

5 9.2

5.1 question

9.2. Fitting a Poisson Distribution: You count the number of times that the annoying "MacSweeper" popup window appears per hour when you surf the web. You wish to model these counts with a Poisson distribution. On day 1, you surf for 4 hours, and see counts of 3, 1, 4, 2 (in hours 1 through 4 respectively). On day 2, you surf for 3 hours, and observe counts of 2, 1, 2. On day 3, you surf for 5 hours, and observe counts of 3, 2, 2, 1, 4. On day 4, you surf for 6 hours, but keep only the count for all six hours, which is 13. You wish to model the intensity in counts per hour.
(a) What is the maximum likelihood estimate of the intensity for each of days 1, 2, and 3 separately?
(b) What is the maximum likelihood estimate of the intensity for day 4?
(c) What is the maximum likelihood estimate of the intensity for all days taken together?

5.2 answer

After reading the section, I found the maximum likelihood formula for the Poisson distribution is

$$\hat{\theta} = \frac{\sum_i n_i}{N}$$

(a)

$$\text{First day:} \frac{\sum_i n_i}{N} = \frac{(3+1+4+2)}{4} = \frac{10}{4} = 2.5 = \hat{\theta}$$

$$\text{Second day:} \frac{\sum_i n_i}{N} = \frac{(2+1+2)}{3} = \frac{5}{3} = 1.6\bar{6} = \hat{\theta}$$

$$\text{Third day:} \frac{\sum_i n_i}{N} = \frac{(3+2+2+1+4)}{5} = \frac{12}{5} = 2.4 = \hat{\theta}$$

(b) Because we are given the total occurrences for the forth day, we can treat that as the sum of all the individual periods. And we know there were 6 periods because that's how many hours were surfed.

$$\text{First day:} \frac{\sum_i n_i}{N} = \frac{13}{6} = \frac{13}{6} = 2.16\bar{6} = \hat{\theta}$$

(c) The formula should still hold over different days because Poisson distribution models independent events.

$$\text{Combined days:} \frac{\sum_i n_i}{N} = \frac{10+5+12+13}{4+3+5+6} = \frac{30}{18} = 1.6\bar{6} = \hat{\theta}$$

6 9.4

6.1 question

9.4. Fitting a Binomial Model: You encounter a deck of Martian playing cards. There are 87 cards in the deck. You cannot read Martian, and so the meaning of the cards is mysterious. However, you notice that some cards are blue, and others are yellow.
(a) You shuffle the deck, and draw one card. It is yellow. What is the maxi- mum likelihood estimate of the fraction of blue cards in the deck?
(b) You repeat the previous exercise 10 times, replacing the card you drew each time before shuffling. You see 7 yellow and 3 blue cards in the deck. What is the maximum likelihood estimate of the fraction of blue cards in the deck?

6.2 answer

(a) So to solve this problem, we will use the ML formula for the binomial to find the probability of getting blue. The formula is $\hat{\theta} = k/N$ where k is the number of times that result appears and N is the number of trials. For this we did one trial, so $N = 1$ and $k = 0$ because we got no blue cards.

$$P(\text{blue card}) = \hat{\theta} = \frac{k}{N} = \frac{0}{1} = 0$$

So the fraction of blue cards in the deck is equal to the probability of getting a blue card, in this case. Therefore according to the maximum likelihood estimate, THERE ARE NO BLUE CARDS IN THE DECK.

(b) Using the same reasoning as a, we can simply plug into the formula the number of blue cards:3 out of the number of trials:10 to get

$$P(\text{blue card}) = \hat{\theta} = \frac{k}{N} = \frac{3}{10} = 0.3$$

$$(\text{Est: Fraction of blue cards}) = .3$$